

Д.Р. Рахимова^{1,2}, Ә.Т. Турарбек^{1,2*}, Р.Р. Галимова³, А.М. Акимбаева⁴

¹Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы қ., Қазақстан

²Ақпараттық және есептеуіш технологиялар институты, Алматы қ., Қазақстан

³Түзеу үлгісіндегі сөйлеу мүмкіндіктері шектеулі балаларға арналған № 54 бөбекжай-балабақшасы, Алматы қ., Қазақстан

⁴№6 Мүмкіншіліктері шектеулі балаларға арналған арнайы мектеп-интернаты, Алматы қ., Қазақстан

*e-mail: turarbek_asem@mail.ru

WORD2VEC ӘДІСІ НЕГІЗІНДЕ СӨЙЛЕУ ҚАБІЛЕТІ НАШАР БАЛАЛАРҒА АРНАЛҒАН ОҚЫТУ ЖҮЙЕСІН ӘЗІРЛЕУ

Аңдатпа

Мектеп жасына дейінгі балалардың сөйлеу қабілетінің нашарлауы қоғамдағы күрделі мәселеге айналды. Арнайы орталықтар мен мамандарға жүгінетін ата-аналардың саны жылдан жылға артып келеді. Зерттеудің мақсаты – баланың дамуында тиімділікке қол жеткізуге ықпал ететін жаңа технологияларды пайдалана отырып, сөйлеуді түзету саласында ақпараттық жүйені құру қажеттілігі. Бұл мәселені шешу үшін қазақ тіліндегі жаңа қосымша әзірленді. Мақалада сөйлеу тілі бұзылған балаларға қазақ тілін оқыту жүйесін құру әдістемесі, соның ішінде қазақ тілінің синонимдер сөздігі қарастырылған. Синонимдер сөздігі баланың деңгейіне қарай сөздік қорын молайта отырып, дұрыс және ана тілінде сөйлеу дағдыларын дамытуға бағытталған. Ұсынылып отырған тәсілдің жаңалығы қазақ тіліндегі мәтіндердегі мағыналық жағынан ұқсас сөздерді анықтауында. Бұл жұмыс логопедия саласындағы жаңа педагогикалық жүйені автоматтандыруға ықпал етеді.

Түйін сөздер: машиналық оқыту, Word2Vec, сөздік, қазақ тілі, логопедия, мобильді қосымша.

Д.Р. Рахимова^{1,2}, А. Т. Турарбек^{1,2}, Р.Р. Галимова³, А.М. Акимбаева⁴

¹Казахский национальный университет имени аль-Фараби, г. Алматы, Казахстан

²Институт информационных и вычислительных технологий, г. Алматы, Казахстан

³Коррекционный ясли сад №54 для детей с тяжелыми нарушениями речи, г. Алматы, Казахстан

⁴Специальная школа-интернат №6, г. Алматы, Казахстан

РАЗРАБОТКА СИСТЕМЫ ОБУЧЕНИЯ ДЕТЕЙ С НАРУШЕНИЯМИ РЕЧИ НА ОСНОВЕ МЕТОДА WORD2VEC

Аннотация

Нарушение речи у детей дошкольного возраста стало серьезной проблемой общества. Из года в год увеличивается количество родителей, обращающихся в специальные центры и специалистов. Целью исследования является необходимость создания информационной системы в области коррекции речи, применяя новые технологии, которые будут способствовать достижению эффективности в развитии ребенка. Чтобы решить эту проблему, мы разработали новое приложение на казахском языке. В статье описан способ построения системы обучения казахскому языку детей с нарушениями речевых способностей включающий словарь синонимов казахского языка. Словарь синонимов работает на развитие речевых навыков правильно и на родном языке, увеличивая словарный запас в зависимости от уровня ребенка. Новизна предлагаемого подхода заключается в выявлении семантически близких по смыслу слов в текстах на казахском языке. Данная работа способствует автоматизации новой педагогической системы в области логопедии.

Ключевые слова: машинное обучение, Word2Vec, словарь, казахский язык, логопедия, мобильное приложение.

D. Rakhimova^{1,2}, A. Turarbek^{1,2}, R. Galimova³, A. Akimbaeva⁴

¹al-Farabi Kazakh National University, Almaty, Kazakhstan

²The Institute of Information and Computational Technologies, Almaty, Kazakhstan

³Correctional nursery garden №54 for children with severe speech disorders, Almaty, Kazakhstan

⁴Special boarding school №6, Almaty, Kazakhstan

DEVELOPMENT SYSTEM LEARNING DETAILS WITH STRUCTURAL LANGUAGES BASED ON WORD2VEC

Abstract

Speech impairment in preschool children has become a serious problem in society. The number of parents turning to special centers and specialists increases from year to year. The purpose of the study is the need to create an information system in the field of speech correction, using new technologies that will contribute to achieving effectiveness in the development of the child. To solve this problem, we have developed a new application in the Kazakh language. The article describes a method for constructing a system for teaching the Kazakh language to children with speech impairments, including a dictionary of synonyms of the Kazakh language. The dictionary of synonyms works to develop speech skills correctly and in the native language, increasing vocabulary depending on the child's level. The novelty of the proposed approach lies in the identification of words that are semantically similar in meaning in texts in the Kazakh language. This work contributes to the automation of a new pedagogical system in the field of speech therapy.

Keywords: machine learning, Word2Vec, dictionary, Kazakh language, speech therapy, mobile application.

Негізгі ережелер

Заман ағымының әсерінен әлемде адамдардың өмір сүру дағдылары мен құндылықтарының өзгеруіне байланысты сөйлеу дамуында ауытқулары бар балалар санының өсуі байқалады. Олардың арасында интеллекті мен есту, көру қабілеті жетілгенімен сөйлеу функциялары нашар балалар тобы ерекшеленеді. Бұл балалардың көпшілігі жалпы білім беретін мектепке баратындықтан, олар оқу және бірқатар қиындықтарға тап болады. Сол себепті сөйлеу қабілеті нашар балалар үшін қазақ тілінде оқыту жүйесін құру өте маңызды.

Кіріспе

Сөйлеу дамуының кешігуін зерттеудің өзектілігі соңғы уақытта осы патологиясы бар балалар санының артуы фактісімен анықталады. Балалық шақта сөйлеуді қалыптастырудың барлық теориялары генетикалық бейімділіктерді жүзеге асыруға ықпал ететін туа біткен қабілеттер мен қоршаған орта факторларының өзара әрекеттесуіне баса назар аударады [1]. Сөйлеу кемшіліктері салдарынан баланың өзіндік қабілеттеріне деген сенімсіздігін тудыруы және сәтсіздікке әкелуі мүмкін. Көп жағдайда балаларда белгілі бір дәрежеде дыбыстық айтылымның ғана емес, сонымен қатар лексикалық құрамның, сөйлеудің грамматикалық құрылымының, фонемалық процестердің бұзылуы болады. Мұның бәрі балаларды мектепке дайындау процесін қиындатады. Қазақстанның өзінде «соңғы 10 жыл ішінде сөйлеу қабілеті нашар балалардың саны 25 пайызға өсті», – дейді түзету педагогикасы ғылыми-практикалық орталығының мамандары. Ресми түрде 45 мыңға жуық сәби есепте тұр. Әсіресе, 1-сынып оқушыларының арасында осындай диагноздағы балалар саны артқанын байқауға болады. Әр сыныпта аз дегенде сөйлеу қабілеті нашар 4-5 оқушы оқиды [2]. Бұл жағдайға көбіне қазақы менталитеттің әсері де тигені анық. Ата-аналардың басым көпшілігі баланың бұндай әрекетін қалыпты санап, алдын ала арнайы балабақшаға берместен оларды мектепке оқытады. Ал баланың жетік сөйлей алмауы бала психикасының кеш дамуынан бастап, жоғарыда атап өткендей түрлі қиындықтарға әкеліп соғады. Сөйлеу қабілетінде нашар балаларға арналған арнайы бағдарламалар мен түрлі іс-шаралар жасалғанмен, гаджет пен телефонда ойнап отырып сөйлеуге талпынбайтын балалар саны өте көп. Сөйлеу қабілеті нашар мектепке дейінгі жастағы балалардың сөйлеу тілін дамытуға арналған қарапайым оқыту құралдарының қазақ тіліндегі нұсқасын табу қиын. Ал оқитын балалардың саны күн санап күрт өсуде.

Қазіргі қоғамды ақпараттық компьютерлік жүйелерсіз елестету қиын. Сөйлеу қабілеті нашар балалар үшін түрлі электронды орта, жаңа технология бала қызығушылығын арттырады. Заманауи ақпараттық технологиялар логопед-мұғалімге логопедтік түзету сабақтарын нәтижелі өткізуге көмектеседі [3,4]. Логопедия саласының даму кезеңінде жаңа педагогикалық үрдіс, фронтальды логопедиялық сабақ өткізуге жаңа әсер беретін заманауи әдістерді қолдану қажеттілікті туындап отыр. Дамыған мемлекеттердің тәжірибелеріне сүйенсек, балаларды сөйлеуге үйретіде инновациялық технологияларды пайдалану сөйлеу қабілеті нашар балалардың пайыздық көрсеткішін азайта алады. Ол үшін баланың сөйлеу қабілетін жақсартуға арналған құралдары әзірлеу, елде болмаған аппараттар мен логопедия орталықтарының ақпараттық жүйесін құру қажет.

Соңғы жылдарда балалардың тілдік дағдыларын жетілдіруге бағытталған, білім беру ресурстарын әзірлеуде табиғи тілді өңдеу Natural language processing (NLP) әдістерін қолдану кеңінен танымал болды. Осындай әдістердің бірі Word2Vec әдісі болып табылады, ол әртүрлі тілдерге арналған синонимдер мен антонимдердің сөздіктерін жасау үшін қолданылған. Зерттеудің басты мақсаты Word2Vec алгоритмін пайдалана отырып, табиғи тілді өңдеу саласында логопедиялық оқыту құралы ретінде синонимдік сөздікті құру және оны, сөйлеуді қабілетін арттыруға арналған жүйеде қолдану болып табылады. Синонимдік сөздік Word2Vec әдісі арқылы тек сөз тіркесін ғана емес, сонымен қатар сөз тіркесінің жақындықтарды да анықтауға мүмкіндік береді.

Мысалы, Kumar зерттеуінде (2019) Word2Vec әдісі хинди тілінде жиі қолданылатын сөздердің синонимдері мен антонимдерін қамтитын хинди тезаурусын әзірлеуде пайдаланды. Сөйлеу қабілеті нашар балалардың тіл байлығын арттыруда тезаурус пайдалы екені анықталды [5]. Сол сияқты, Mavridis (2020) Word2Vec әдісін балалар жиі қолданатын сөздердің синонимдері мен антонимдерін қамтитын грек тезаурусын жасау үшін пайдаланылды [6].

Қазақ тілі контекстінде Серікболова мен Шөкеева (2019) Word2Vec әдісі арқылы қазақ тілінің синонимдер сөздігін жасады. Алынған сөздік 500-ден астам сөздер мен сөз тіркестерін қамтиды, олар тақырыптық түрде ұйымдастырылған және суреттер мен қарапайым анықтамалармен сүйемелденеді. Сөздіктің сөйлеу қабілеті нашар балалардың тіл байлығын арттыруға пайдасы бар екені анықталды [7]. Серікболова мен Шөкеева (2020) бұрынғы жұмыстарына сүйене отырып, қазақ тілінің синонимдер сөздігінің кеңейтілген нұсқасын әзірледі. Сөздік 800-ден астам сөздер мен сөз тіркестерін қамтиды, сонымен қатар тағам, жануарлар және эмоциялар сияқты категориялар бойынша тақырыптық түрде ұйымдастырылған. Сөздік Қазақстандағы мектеп жасына дейінгі балалардың тіл байлығын арттыруда пайдалы екені анықталды [8]. Серікболова мен Шөкеева (2021) өздерінің соңғы зерттеулерінде қазақ тілінің синонимдер сөздігін 1000-нан астам сөз бен сөз тіркесіне кеңейтті. Сөздік сөйлеу қабілеті нашар мектеп жасына дейінгі балаларға арналған және оқу орындарында пайдалануға арналған. Сөздік түстер, пішіндер, тұрмыстық заттар сияқты категориялар бойынша тақырыптық түрде ұйымдастырылған, суреттер және қарапайым анықтамаларды қамтиды. Зерттеу барысында сөздіктің Қазақстандағы мектеп жасына дейінгі балалардың тіл байлығын арттыруда тиімді екені анықталды [9].

Сонымен қатар, осы негізінде қазақ тіліндегі синонимдерді анықтауға және қазақ тілін оқытуға бағытталған онлайн платформа жұмыс істейді. Бұл әр түрлі салалық сөздіктер мен энциклопедиялардағы сөздер мен тұрақты тіркестердің, қазақ тіліндегі көне сөздердің, кірістірме сөздердің, аймақтық және ақпараттық технологиялардың даму кезеңінде жаңа технологиялық сөздердің мағынасын көруге мүмкіндік беретін алаң. Сөздік порталының іздеу жүйесі арқылы фразеологиялық тіркестегі немесе сөйлем ішіндегі сөздердің, синонимдердің, антонимдердің, омонимдердің, кездесулердің анықтамасын бір бетте көруге болады. Қазіргі уақытта қорда 1 243 850 сөз бар. Платформа кез келген сөздің синонимін табуға көмектеседі.

Бұл зерттеулер Word2Vec әдісінің әртүрлі тілдерге, соның ішінде қазақ тіліне арналған синонимдік сөздіктерді жасаудағы тиімділігін анықтайды және сөйлеу қабілеті нашар

балалардың тілдік дағдыларын жетілдіруді әдісі екенін көрсетеді. Зерттеулер мектеп жасына дейінгі балалар үшін ресурстарды қол жетімді және ыңғайлы ету үшін сөздіктерді тақырыптық түрде ұйымдастырудың және арнайы суреттер мен қарапайым анықтамаларды берудің маңыздылығын ұсынады. Сөйлеу қабілеті нашар балалардың саны күн сайын артып келеді. 2021 жылдың тамызына [10] жаңа оқу жылына Алматының өзінде сөйлеу қабілеті нашар 640 мектеп жасына дейінгі бала тіркелді. Осындай өзекті мәселені шешу үшін табиғи тіл өңдеуді қолданып, кем дегенде баланың сөздік қорын дамытатын синонимдік сөздік жасауға болады. Өзірленген әдіс баланың жасы мен оқу қабілетін ескеруге мүмкіндік береді. Жасалған синонимдік сөздік сөйлеу қабілеті нашар балаға оқуға дұрыс тақырыптық сөздерді құрастыруға мүмкіндік береді. Бұл тез ассимиляцияға және сөздік қорын кеңейтуге мүмкіндік береді. Логопедия саласында мұндай оқу құралдары жоқтың қасы. Бұл зерттеу жұмысының мақсатты аудиториясы – сөйлеу дағдылары нашар мектеп жасына дейінгі балалар мен логопедия саласындағы кәсіби мамандар.

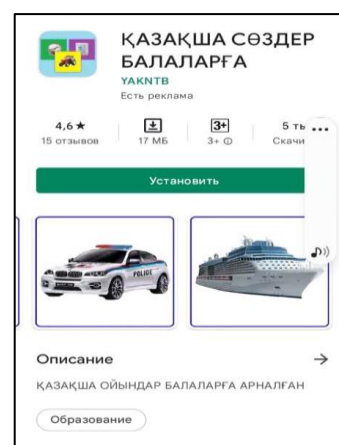
Зерттеу әдіснамасы

Ақпараттық технологияларды логопедия саласында қолдану сөйлеу қабілеті нашар балалар үшін өте маңызды болып келеді. Ал оны қазақ тілінде сөйлеуді үйрету және қазақ тілінде сөйлетін балалар үшін оны қолдану өте қажет. Қазақ аудиториясына арналған қосымшалар мен құралдарға талдау жасайтын болсақ, Android жүйесіне арналған жалғыз ғана қосымша бар. Оның өзі 1-суретте көрсетілгендей арнайы логопед маманның курсына арналған Майра Болатбекқызының тапсырысымен жасалған ақылы қосымша. Қосымша біріншіден, ақылы болғандықтан көп адамға қолжетімсіз. Екіншіден, екі бөлімнен ғана тұрады, олар – ойындар және видео, аудио сабақтар. Қосымшаның сипаттамасына қарағанда үй жағдайында баланың онлайн түрде дайындығын қамтамасыз етеді.

Бұдан бөлек, 2-суретте көрсетілгендей қазақ тіліндегі балаларға сөздерді айтып үйретуге арналған ойындар топтамалары бар.



Сурет 1. «Өз-өзіне Логопед!» ең алғашқы қазақ тіліндегі қосымша



Сурет 2. «Қазақша сөздер балаларға» Android қосымшасы

Технология дамыған заманда кез келген ақпараттың тез әрі сапалы, қолжетімді жеткізудің жолын қарастыру шарт. Орыс тілінде ақпарат қарапайым халыққа қолжетімді болып келеді. Оның айқын дәлелі, осы логопедия саласынан орын алып отыр. Қосымша тегін, қолдануға ыңғайлы жасалғанымен, дизайны мен қамтитын ақпараттық ресурсы аз. Логопедия саласында айтарлықтай бәсекелестіктің жоқ екеніне осы екі қосымша дәлел. Ендігі мақсат логопедия саласында ақпараттық жүйе мен технологияны дамыту. Балаларға арналған қазақ тіліндегі жаңа заман ақпараты көбейген сайын қазақ тілінің мазмұнының кеңеюіне үлес қосу керек (1-кесте).

Кесте 1. Орыс және ағылшын тіліндегі мобильді қосымшаларды пайдаланушы тұрғысынан талдау

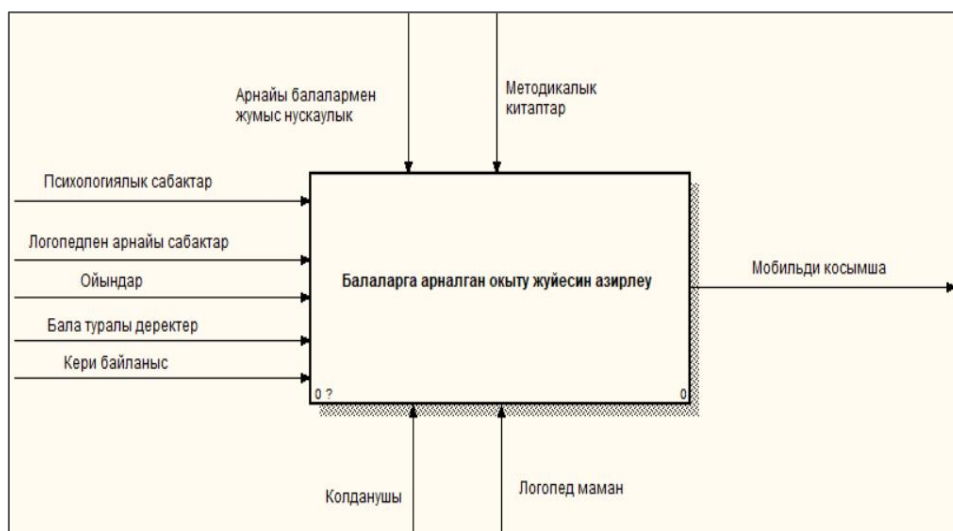
Қосымшаның атауы	Орыс тіліндегі мобильді қосымшалар			Ағылшын тіліндегі мобильді қосымшалар		
	«Домашний логопед»	«ЛОГО-ПЕДУЯ»	«Развитие речи у детей»	«Speech Blubs: Language Therapy»	«Speech Therapy: Practice S»	«Therapy for Autism & Speech & Learning Disabilities»
Категориялар						
Ыңғайлы интерфейс	+	+	+	+	+	+
Видео, аудио форматтағы файлдар	+	+	+	+	+	+
Ойындар	+	-	+	+	+	+
Ақысыз бағдарлама	+	+	+	+	+	+
Кері байланыс	-	-	-	-	+	+

Талдау нәтижесін қорытындыласақ, ағылшын тілді бағдарламаларға қарағанда орыс тілді аудиторияға арналған мобильді қосымшалар қатары қарқынды дамыған. Кез келген форматта тегін немесе ақылы түрде жүктеуге болады. Ағылшын тіліндегі қосымшалар жүйелендірілген, яғни аудиториясы нақты, ұсынатын ақпараты анық және интерфейсі ыңғайлы ойластырылған. Ал орыс тілінде құрылған қосымшаларда жүйесіз жасалған, қолдануға ыңғайсыз, қамтитын ақпарат ауқымы аз қосымшалар. Кері байланыс функциясы қарастырылған қосымша міндетті түрде ақылы негізде ұсынылады. Сонымен қатар, мобильді қосымшаны байқап көру үшін, бастапқы деңгейлерін тегін қолжетімді етіп, қалған ең керек деген ақпараттар мен функцияларды ақылы түрде жасаған. Егер қосымша толық тегін әрі маманмен кері байланысы болса, ол маманның қызметін сатуына негізделген болып табылады. Осылайша жүйелерге жасалған талдау негізінде қазақ тілінде сөйлеуді үйрететін мобильді қосымшаны құру үшін көптеген мәселелер ескерілу қажет. Балаларға арналған оқытудың ақпараттық жүйесін, мобильді қосымша құралы ретінде құрастыру процесін модельденді. Модельде жүйенің кіріс, шығыс деректері көрсетіледі. Оларға:

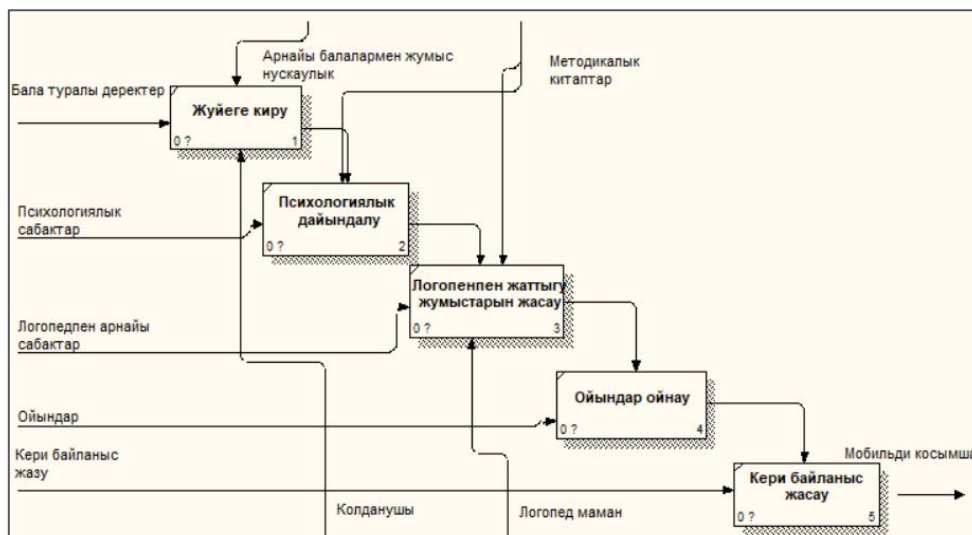
- Кіріс: психологиялық сабақтар, логопедпен арнайы сабақтар, ойындар, бала туралы деректер және кері байланыс;
- Шығыс: жүйе дайын ақпараттық мобильді қосымша негізінде құрылып дайындалады;
- Басқарушы: арнайы балалармен жұмыс нұсқаулығы және әдістемелік кітаптар;
- Механизмдер: логопед маман мен қолданушы.

Таңдалған ақпаратпен 3-суретте көрсетілгендей IDEF0 AS-IS моделінің А0 блогында модель сыздық. 4-суретте көрсетілген балаларға арналған оқыту жүйесін әзірлеудің негізгі жұмысын бес тапсырмасы бөлініп, бейнеленген:

- жүйеге кіру;
- психологиялық дайындалу;
- логопедпен арнайы жұмыс жасау;
- ойындар ойнау;
- кері байланыс жасау.



Сурет 3. Балаларға арналған оқыту жүйесін әзірлеу процесінің IDEF0 AS-IS моделі. А0 Блогы



Сурет 4. Балаларға арналған оқыту жүйесін әзірлеу процесінің IDEF0 AS-IS моделі. А0 блогының декомпозициялануы

Жүйе компонентті іске қосқан кезде, компонентке тиесілі бағдарлама процесі басталады және компонентке қажет класс даналарын құрады. Сондықтан, көптеген басқа жүйелерден айырмашылығы, Android жүйесінде қосымшада бірыңғай кіру нүктесі жоқ. Әр қосымшаның жеке процесте іске қосылуына және файлдарға кіру шектеулеріне байланысты бағдарлама басқа компонентті тікелей іске қоса алмайды. Осылайша, ол үшін жүйеге белгілі бір компонентті іске қосу туралы хабарлама жіберу керек, жүйе оны іске қосады.

Компоненттердің төрт түрі болады, олардың әрқайсысы белгілі бір мақсатқа жетуге қызмет етеді және тиісті компонентті құру және жою тәсілдерін анықтайтын өзіндік циклі болады. Android қосымшаларының негізгі компоненттерін қарастырылды, олар:

1. Әрекет (Activities) – бұл қосымшаның көрінетін бөлігі (экран, терезе, пішін), графикалық пайдаланушы интерфейсіні көрсетуге жауап береді;

2. Қызмет (Services) – фондық режимде, ұзақ уақыт жұмыс істейтін немесе қашықтағы процестер үшін жұмыс жасайтын компонент;

3. Контент-провайдерлер (Content providers) – мазмұн провайдері қосымша деректерінің таратылған жиынтығын басқарады;

4. Хабар тарату қабылдағыштары (Broadcast Receivers) – қабылдағыш-хабар тарату хабарламаларына жауап беретін компонент [11].

Жалпы, Android қосымшасы мынадай бөліктерден тұрады:

- Android SDK (View, Activity, ContentProvider, Service, Broadcast-Receiver, Intent) негізгі кластары болып табылатын Java кластары;

- қосымшаның манифестасы;

- жолдар, суреттер және т. б. сияқты ресурстар;

- файлдар.

Қосымша ішіндегі деректер Firebase реляциялық деректер қоры арқылы өңделеді. JSON деректерін Firebase-де сақтаған кезде, өзгерістер оларды сұраған барлық клиенттер, веб және мобильді құрылғылар үшін бірден жібереді. Кірістірілген тұрақты файл хостинг, пайдаланушы басқару және қауіпсіздік ережелерімен Firebase заманауи қосымшаларды бұрынғыдан тезірек құруға көмектеседі. Firebase деректердегі өзгерістер туралы хабарлау үшін, сондай-ақ алғаш деректерді алу үшін оқиғаларға негізделген модель қолданылады [12].

Firebase деректер қорының мүмкіндіктері:

- Analytics – қосымша бойынша аналитика: аудитория мөлшері, пайдаланушылар туралы ақпарат, қосымшадағы оқиғалар және т.б.;

- Authentication – пайдаланушылар өздерінің есептік жазбаларын қосымшаға байланыстыра алады және оларға кез-келген деректер қосылады. Бұл жерден келесі провайдерлерге қолдау көрсетіледі: Google, Facebook, Twitter, GitHub;

- Realtime Database – тіркеу үшін нақты уақытта өзгерістермен жұмыс істейтін деректер қоры.

Құрылған қосымшаның басты ерекшелігі «Логопедпен жұмыс жасау» бөлімінде синонимдер сөздігі кірістірілген. Қазақ тіліндегі сөйлеу қабілеті нашар мектеп жасына дейінгі балаларға арналған синонимдер сөздігін құру үшін табиғи тілі өңдеу әдісі ретінде Word2Vec әдісі таңдалды. Бұл әдіс контекст негізінде семантикалық және синтаксистік мағынаны қамтитын сөздерді енгізу үшін кеңінен қолданылды. Ол терезе өлшемі, вектор өлшемі және ең аз сөз жиілігі сияқты пайдаланылған гиперпараметрлерді анықтайды. Сондай-ақ Word2Vec әдісімен оқыту үшін пайдаланылатын дәуірлер саны мен пакет өлшемін қоса алғанда, оқыту процесін талқылайды. Онда модельдің өнімділігін бағалау және сарапшылардан алынған кері байланыс туралы айтылады. Сөздерді кірістіру мағынасы ұқсас сөздерді анықтауға және синонимдер сөздігін жасауда сәйкес баламалар беруге қолайлы тәсілді қамтамасыз етеді [13].

Корпусты алдын ала өңдегеннен кейін теріс сынамамен Skip-gram архитектурасын пайдаланып, Word2Vec әдісі оқытылды. Тренинг Python тіліндегі Gensim кітапханасы арқылы келесі гиперпараметрлермен орындалады:

- Терезе өлшемі: 5;

- Вектор өлшемі: 100;

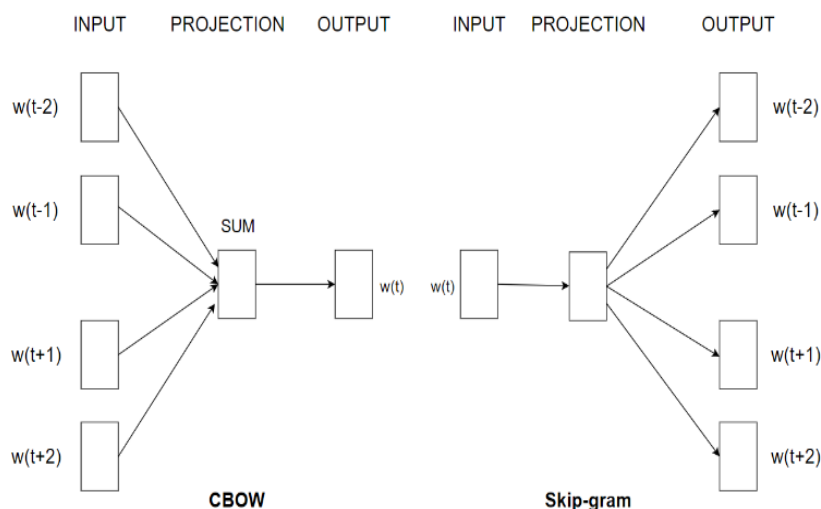
- Минималды сөз жиілігі: 5.

Терезе өлшемі параметрі оқытуда қарастырылатын мақсатты сөз бен контекстік сөздер арасындағы ең үлкен қашықтықты анықтайды. Терезенің кішірек өлшемі сөздер арасындағы жергілікті қарым-қатынастарды түсіруге бейім, ал үлкен терезе өлшемі аумақты қатынастарды түсіреді. Бұл зерттеуде тақырып бойынша алдыңғы зерттеулер негізінде 5 терезе өлшемін таңдалған.

Word2Vec әдісін таңдау бірнеше факторларға негізделді, оның ішінде бақылаусыз оқыту сипаты, үлкен көлемдегі деректерді тиімді өңдеу мүмкіндігі және сөз енгізу тиімділігі. Бұл әдіс оқыту үшін қолмен белгіленген деректерді қажет етпейді, бұл оны қазақ тіліне қолайлы етеді. Жиналған корпус алдын ала тоқтау сөздерін, тыныс белгілерін және басқа маңызды емес деректерді алып тастау арқылы өңделді, содан кейін олардың өлшемдерін азайту үшін қалған сөздерді лексемалау және түбірлеу жүргізілді [14].

Word2Vec әдісінің екі алгоритмі бар, олар 5 суретте көрсетілген:

- CBOW;
- Skip-gram.



Сурет 5. Екі Word2Vec моделінің салыстырмалы алгоритмі [15]

CBOW. Үздіксіз сөмке (CBOW) Word2Vec әдісінің архитектураларының бірі болып табылады, ол контекст негізінде сөздердің векторлық бейнелерін жасау үшін қолданылады. CBOW негізгі идеясы - оның айналасындағы сөздердің контекстіне негізделген ағымдағы сөзді болжау. Мысалы, «Мен ...(кешкі ас) ішкенді ұнатамын» сөйлемінде контекстіне сүйене отырып, «кешкі ас» сөзін болжауға тырысады. CBOW үлгісін үйрету үшін бір жасырын қабаты бар нейрондық желі қолданылады. Кіріс мәтінмәндегі сөздердің векторлары, ал шығыс - біз болжауға тырысып жатқан сөздің векторы. Жасырын қабат нейрондары контекстік сөздерді біріктіреді және сөздіктегі әрбір сөздің ықтималдығын есептейді, содан кейін олар желінің салмақтарын жаңарту үшін пайдаланылады. CBOW артықшылықтары қысқа мәтіндер мен бірнеше бірегей сөздері бар мәтіндерде жақсы жұмыс істеуді қамтиды. CBOW қарапайым мәтіндерді Skip-gramға қарағанда жылдамырақ үйренеді. Дегенмен, құрылымы күрделірек мәтіндерде немесе бірегей сөздердің көп саны бар мәтіндерде CBOW дәлдігі азырақ. Сондай-ақ, CBOW әдетте сөз векторларын сақтау үшін көбірек жады қажет етеді.

$$Q=N \times D+D \times \log _2(V) \quad (1)$$

Мұндағы, Q - параметрлер саны, N - сөзді көрсету векторларының өлшемі, D - контекстік терезе өлшемі, V - сөздіктің өлшемі. Сөзді көрсету векторларының өлшемі («эмбеддинг» деп те аталады) мәтіндегі әрбір сөзді көрсету үшін пайдаланылатын сандық мәндердің саны болып табылады. Мысалы, сөзді көрсету векторының өлшемі 100 болса, онда әрбір сөз 100 санның векторы ретінде көрсетіледі.

Skip-gram. Skip-gram сөздердің векторлық көріністерін жасау үшін қолданылатын Word2Vec моделінің архитектурасы. CBOW-тен айырмашылығы, Skip-gram берілген мақсатты сөз үшін контекстік сөздерді болжауға тырысады. Яғни, егер бізде «Мен ...(кешкі ас) ішкенді ұнатамын» деген сөйлем болса, Skip-gram «кешкі ас» мақсатты сөзіне сүйене отырып, «мен», «ұнату», «ішу» деген сөздерді болжауға тырысады. Skip-gram әрбір ағымдағы сөзді үздіксіз проекциялық қабаты бар сызықты классификаторға кіріс ретінде қолданады және ағымдағы сөзге дейін және кейін белгілі бір аралықтағы сөздерді болжайды. Аралықты ұлғайту нәтижесінде алынған сөз векторларының сапасын жақсартатынын анықтадық, бірақ ол сонымен қатар есептеу күрделілігін арттырады. Бұл архитектураның оқыту күрделілігі

пропорционалды болады. Нейрондық желі Skip-gram моделін үйрету үшін де пайдаланылады, бірақ CBOW-мен салыстырғанда кері архитектурасы бар. Кіріс - мақсатты сөз және шығыс - контекстік сөздер. Жасырын қабаттағы нейрондар мақсатты сөзді векторға түрлендіреді және оны контекстік терезеде контекстік сөздерді болжау үшін пайдаланады.

Skip-gram артықшылығы сирек сөздерді жақсырақ өңдеуді және сөздер арасындағы семантикалық қатынастарды дәлірек модельдеуді қамтиды. Сондай-ақ Skip-gram мәтіннің үлкен көлемін жақсы меңгереді және күрделі мәтін құрылымдарын дәлірек есептей алады. Дегенмен, Skip-gram көбірек есептеу ресурстары мен оқу уақытын қажет етеді, әсіресе мәтіннің үлкен көлеміне.

$$Q=C \times (D+D \times \log_2(V)) \quad (2)$$

мұндағы C - сөздердің максималды қашықтығы. Осылайша, егер $C=5$ таңдасақ, әрбір жаттығу сөзі үшін кездейсоқ R санын <1 аралығында таңдаймыз; $C >$, содан кейін дұрыс белгілер ретінде тарихтағы R сөздерін және ағымдағы сөздің болашақтағы R сөздерін пайдалану керек. Бұл ағымдағы сөзді кіріс ретінде және $R+R$ сөздерінің әрқайсысын шығыс ретінде көрсете отырып, $R \times 2$ сөз жүктелуін талап етеді. Келесі тәжірибеде $C=10$ [16] қолданылған. CBOW және Skip-gram екеуінің де артықшылықтары мен кемшіліктері бар. CBOW тезірек жаттығады және жиі сөздерді жақсы орындауға бейім, ал Skip-gram жаттықтыруда баяу, бірақ сирек сөздерді жақсырақ орындауға бейім және сөз қатынастары туралы толық ақпаратты жинайды [17].

Зерттеу нәтижелері

Бұл зерттеуде Word2Vec әдісі негізінде қазақ тілінде сөйлеу қабілеті нашар мектеп жасына дейінгі балаларға арналған синонимдер сөздігінен тұратын мобильді қосымша жасауды мақсат етілді. Ол үшін қазақша мәтіннің корпусын жиналды және оны тоқтау сөздерін, тыныс белгілерін және басқа да маңызды емес деректерді алып тастау үшін алдын ала өңделді, содан кейін қалған сөздерді лексемалау және түбірлеу жүргізілді. Word2Vec әдісі корпустағы әрбір сөз үшін сөз енгізу алдын ала өңделген корпуста оқытылады. Келесі қадам ретінде оқыту жүйесінің моделін құрылды, оның ішіне жоғарыда сипатталған синонимдер сөздігі қосылды. Word2Vec әдісінің өнімділігін бағалау үшін екі көрсеткіш қолданылды: косинус ұқсастығы және сөз ұқсастығы. Косинус ұқсастығы екі вектор арасындағы ұқсастық дәрежесін өлшейді, ал сөз ұқсастығы екі сөз арасындағы ұқсастықты кірістіру негізінде есептейді. Сөздердің жұптары арасындағы косинус ұқсастығы мен сөз ұқсастығын есептеп, нәтижелерді адам белгілеген ұқсастықпен салыстырылды [18].

Gensim. Gensim – табиғи тілді өңдеуге арналған танымал Python кітапханасы, ол сөз ендірілген сөздерді құру және пайдалану құралдарын ұсынады. Оның ең танымал мүмкіндіктерінің бірі Word2Vec әдісі болып табылады, ол мәтін корпусындағы бірлесе пайда болу статистикасына негізделген жоғары сапалы сөздерді енгізу үшін қолданылады. Gensim сөзді ендіруден басқа, сонымен қатар тақырыпты модельдеуге, ұқсастық сұрауларына және басқа табиғи тілді өңдеу есептеріне арналған құралдарды ұсынады.

Деректерді жинау және алдын ала өңдеу. Бұл зерттеуде пайдаланылған қазақ тілінің деректер жинағы балалардың әдеби кітаптарынан және әлеуметтік желілерден жиналды. Деректер жинағы шамамен 300 мың сөзден тұрды. Содан кейін таңбалау, кіші әріппен жазу және тоқтату сөздері мен тыныс белгілерін алып тастау сияқты алдын ала өңдеу қадамдары орындалды. Word2Vec әдісін үйрету үшін пайдаланылған соңғы деректер жинағы шамамен 70 мың сөзден тұрды. Корпус деректері бірнеше негізгі тақырыптар бойынша жіктелді: объектілер, табиғат, жануарлар, отбасы.

Модельдік тренинг. Енгізу өлшемі 100, 4 жұмысшы және ең аз саны 1 болатын Gensim кітапханасын пайдаланып Word2Vec әдісі оқытылды. Модель алдын ала өңделген деректер жиынында оқытылды және оқу процесі 16 Гб жедел жады бар құрылғыда шамамен 30 минутқа созылды.

Бағалау көрсеткіштері. Word2Vec әдісі екі метрика арқылы бағаланды: (1) косинус ұқсастығы және (2) ең ұқсас сөздер. Енгізу кеңістігіндегі екі сөздің ұқсастығын өлшеу үшін косинус ұқсастығы қолданылды [19]. Ұқсас сөздердің көпшілігі берілген сөздің синонимдерін анықтау үшін қолданылған.

Синонимді алу. Берілген сөздің синонимдерін шығару үшін Word2Vec әдісімен берілген ең ұқсас сөздер қызметі қолданылды. Мысалы, «жеміс» сөзінің синонимдерін шығару үшін келесі код қолданылды:

```
#Embedding size
Embedding_Dim = 100
#train word2Vec model
model = gensim.models.Word2Vec(sentences = final, size = Embedding_Dim, workers = 4,
min_count = 1)
word1 = 'Алма'
word2 = 'жеміс'
Vector1 = model.wv[word1]
Vector2 = model.wv[word2]
similarity = np.dot(Vector1, Vector2) / (np.linalg.norm(Vector1) * np.linalg.norm(Vector2))
print('Cosine similarity between', word1, 'and', word2, ':', similarity)
model.wv.most_similar('жеміс')[:5]
```

Бұл кодтың нәтижесі «жеміс» сөзіне ең ұқсас бес сөзді берді, олар «жеміс», «алма», «алмұрт», «алхоры» және «өрік» болды.

Косинус ұқсастығын бағалау. Косинус ұқсастық метрикасын бағалау үшін екі сөз арасындағы ұқсастықты өлшеп, нәтижелерді олардың белгілі ұқсастығымен салыстырылды. Мысалы, «алма» және «жеміс» арасындағы косинус ұқсастығын өлшеу үшін келесі код қолданылды:

```
Vector1 = model.wv['алма']
Vector2 = model.wv['жеміс']
similarity = np.dot(Vector1, Vector2) / (np.linalg.norm(Vector1) * np.linalg.norm(Vector2))
print('Cosine similarity between', 'алма', 'and', 'жеміс', ':', similarity)
```

Бұл кодтың нәтижесі 0,53 болатын «алма» және «жеміс» арасындағы косинус ұқсастық балл берді. Содан кейін бұл балл олардың белгілі ұқсастығымен 0,68 салыстырып, Word2Vec моделінің олардың семантикалық ұқсастығын анықтауда жақсы жұмыс істейтінін анықталды, ол 2 кестеде көрсетілген.

Кесте 2. Косинус ұқсастығын бағалау

Түпнұсқа сөз	Косинус ұқсастығы	Адам таңбаланған ұқсастық
Алма – жеміс Apple – fruit	0.53	0.68
Алма – ағаш Apple – tree	0.61	0.71
Жеміс – ағаш Fruit – tree	0.44	0.5
Алмұрт – ағаш Pear - tree	0.6	0.67

Ол ұсынылған әдістің өнімділігін бағалау үшін әртүрлі сөздер жұптары арасындағы косинус ұқсастығы оқытылған Word2Vec әдісін қолданылады. Сонымен қатар, ұсынылған әдіс

сөйлеу қабілеті нашар балаларға арналған қазақ тілінің синонимдерінің сөздігін жасау үшін пайдаланылды. Сөздік деректер жиынындағы әрбір сөз үшін ең ұқсас 5 сөзді табу арқылы құрылады. Нәтижесінде сөздікте әр сөздің көп синонимі бар екенін анықтадық, бұл сөйлеу қабілеті нашар балаларға сөздік қорын және әртүрлі дағдыларын жетілдіруге көмектесуі мүмкін.

Осылайша, Word2Vec арқылы әзірленген қазақ тілінің синонимдер сөздігі сөйлеу қабілеті нашар балаларға арналған сөздердің синонимдерін анықтауда тиімді болып шықты. Косинус ұқсастығын бағалау Word2Vec әдісінің сөздер арасындағы мағыналық ұқсастықты, ал ең ұқсас сөздер қызметі синонимдерді тиімді анықтауға қабілетті екенін көрсетті. Word2Vec әдісі ұқсас мағыналары бар сөздердің жұптары арасындағы жоғары косинус ұқсастыққа қол жеткізді, бұл жасалған кірістіру сөздердің семантикалық мағынасын дәл түсіргенін көрсетеді. Модель мағыналары ұқсас сөздерді дәл анықтай алғанын көрсететін сөздердің ұқсастығының жоғары бағасына қол жеткізді.

Өзара әрекеттесуді жобалау және визуалды дизайн принциптерінің көпшілігі белгілі бір платформамен байланысты емес. Дегенмен, мобильді құрылғылар әртүрлі шектеулерге (экран өлшемі, енгізу әдістері және т.б.) байланысты ерекше ойларды қажет етеді.

Мобильді қосымша атауы «Ainalayn». Сыртқы интерфейсіне пайдаланушыға ыңғайлы болуы үшін барынша қарапайым әрі сәнді үйлесімділікті басты назарға ұсталған. Интерфейс дизайнындағы түстер бала психологиясын ескере отырып таңдалынды. Ашық, айқын, баланы жалықтырмайтын түс болуы бала психологиясына жағымды әсер береді. Интерфейсті мобильді жүйенің алғашқы бетінен-ақ сипаттауға болады (6-сурет). Алғашқы бетте жүйеге кіру және тіркелу батырмалары орналасқан. Қосымша құрылысына пайдаланушы тұрғысынан тоқталатын болсақ, жүйе:

- жүйеге кіру және тіркелу;
- пайдаланушының жеке кабинеті;
- логопедия бөлімі;
- психологиялық бөлім;
- ойындар бөлімі;
- кері байланыс бөліктерінен тұрады.

Жалпы жүйе кімдерге арналған:

- логопедия саласындағы мамандарға;
- сөйлеу қабілетінде қандай да бір кемшілігі бар балаларға;
- осы салада онлайн форматта оқыту жүйесін қолданылатын мамандарға.

Жаңа пайдаланушыны тіркеу тікелей деректер қорымен байланысты. DB_Functions ішінде файлда пайдаланушыны дерекқорға сақтау, пайдаланушыны дерекқордан алу қызметі бар. Сондай-ақ, пайдаланушыны жаңарту, пайдаланушыны жою сияқты әдістерді қосуға болады. Пайдаланушы идентификаторы 6-суретте қосымшаға жаңадан тіркелу беті көрсетілген.

1. Жаңа мәліметтерді енгізген кезде бағдарлама оларды Firebase МҚБЖ (мәліметтер қорын басқару жүйесі) жібереді.

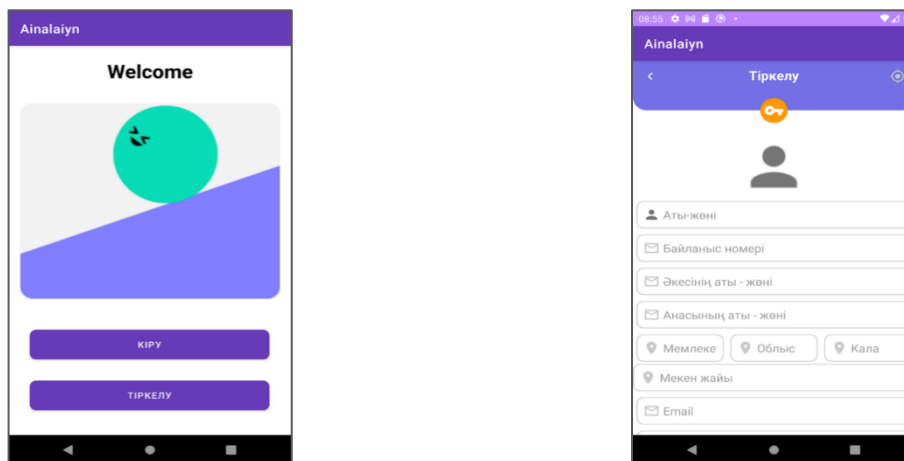
2. Деректер қорында деректерді тексереді және егер олар дұрыс болса, таңбаларды белгілейді, ол қосымшаның кейінгі қосылымдары үшін аутентификацияның дәлелі болып табылады.

3. Кейін мәліметтер бұл таңбаларды қосымшаға жібереді.

4. Бағдарлама таңбаларды сақтайды.

5. Кейінгі авторизация кезінде қосымша қол жетімді белгіні алады және тіркеледі. Құпия сөздер ешқашан қосымшаның жадында сақталмайды.

Тіркелу бөліміндегі мәліметтер атрибуттарға меншіктеліп, пайдаланушының жеке кабинетінде автоматты түрде толтырылатын болады.



Сурет 6. Қосымшаның алғашқы беті. Тіркелу және кіру функциялары

Жаңадан тіркелген пайдаланушы жүйеге кіруі үшін алдымен, енгізілген ақпарат деректер қорымен сәйкестігі анықталатын болады. Деректер қорында жаңа жазба қосылғанда Email and Password() көмегімен пайдаланушыны құру қызметі қолданылады. Бұл қызмет екі міндетті параметрді қабылдайды: электрондық пошта және пароль. Қызмет пайдаланушы сәтті қосылған кезде, сондай-ақ, пайдаланушы сәтсіз қосылған кезде жұмыс істейтін оқиға өңдегіштерін қосылады. Деректер қорынан деректерді алу және пайдаланушыны авторизациялау үшін signInWithEmailAndPassword() әдісін пайдаланылды. Бұл әдіс екі параметрді (email және пароль) қабылдайды және сәтті немесе сәтсіз тіркелу үшін оқиға өңдеуді қамтиды.

7-суретте пайдаланушының жеке кабинеті көрсетілген, ол қосымшаның негізгі беті болып саналады. Алғаш қосымша ашылғанда пайдаланушының жеке кабинеті ашылады. Жеке кабинеттің мәліметтері пайдаланушы тіркелгенде автоматты түрде осы жаққа толтырылады. Негізгі бет ретінде пайдаланушыға қолайлы болуы үшін мәзір құралған.

Android мәзірдің екі түрін қолдайды: негізгі және контексті [20]. Ескі телефондарда жеке мәзір батырмасы бар, оны басу негізгі мәзірді ашады. Смартфондарда оның орнына қосымшаларда жиі қолданылатын ActionBar әрекеттер жолы пайда болды, ол арқылы мәзір оңай құрылады.

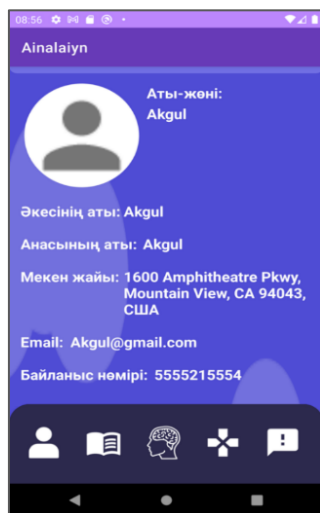
Мәзір құру үшін RES/menu жобасының каталогында XML файлын құрылады және келесі элементтер пайдаланылады:

– <menu> мәзір элементтері үшін контейнер болып табылатын Menu анықтайды. <Menu > элементі файлдың түпкі түйіні болуы тиіс және бір немесе бірнеше <item>және <group> элементтерін қамтуы мүмкін;

– <item> бұл MenuItem мәзірінде бір элемент атауын білдіреді. Бұл элементте ішкі мәзірді құру үшін кірістірілген <menu> элементі болуы мүмкін;

– <group> ол Item элементтері үшін қосымша көрінбейтін контейнер. Бұл мәзір элементтерін белсенді күйі мен көріну сияқты ортақ қасиеттерге ие жіктеу мүмкіндігін қарастырады.

«Ainalayn» қосымшасының пайдаланушыға беретін пайдасы онлайн түрде тегін видеосабақтар. Көпшілікке қолжетімсіз білімді қосымша арқылы ала алады. Android бағдарламасы бейне файлдардың ең көп таралған 3GPP (.3gp), WebM (.webm), Matroska (.mkv), MPEG-4 (.mp4) форматтағы түрлерін қолдайды. Бейне материалдармен жұмыс істеу үшін Android виджеттерінің стандартты жиынтығында бейнені ойнатуға мүмкіндік беретін бейне шолу класы анықталған.



Сурет 7. Пайдаланушының жеке кабинеті

Бейнені қарау мобильді құрылғыда орналастырылған роликтермен де, желідегі бейне материалдармен де жұмыс істей алады. Осылай бейне материалдарды пайдалана отырып, 8-суретте көрсетілген жаңа бөлім логопед маманмен арнайы сабақтарды көре аламыз.

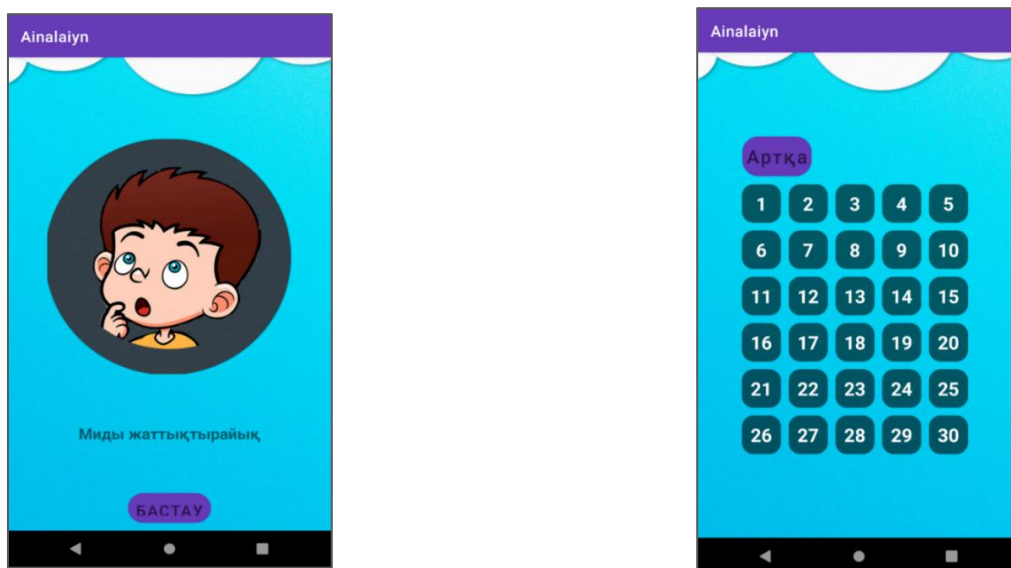


Сурет 8. Логопедпен жұмыс бөлімі. Арнайы бейне сабақтар

Мобильді қосымшаның психологиялық бөлімі пайдаланушының ата-анасымен бірге қолдануға арналған бөлім. Сөйлеу қабілеті нашар балалардың ата-аналарына психологиялық дайындық ретінде, баламен қалай жұмыс жасағанда нәтижелі болатындығы, тағы да көптеген психологиялық білімдер және осы мобильді қосымшаны нәтижелі пайдалануға нұсқаулық ретінде арнайы сабақтар қарастырылған. Осы бөлімдегі видео форматтағы файлдарды алдын ала деректер қорына жүктеп, сілтеме арқылы қолданылады. Сілтеме арқылы қолдану ыңғайлы, уақытты және жады көлемін үнемдеуге көмектеседі. PutBytes() әдісі – файлды деректер

қорына жүктеудің ең тиімді жолы. Жадты аз пайдалану үшін `putStream()` немесе `putFile()` қолдануды қарастырған жөн. Пайдаланушы алған білімді кері байланыс жасауы үшін `.docx` форматтағы файл түрінде ақпараттар ағынын қалдыра алады. Пайдаланушыға пайдалы ақпараттардан бөлек қызықты ойындар ұсыну арқылы сөйлеуге деген құлшынысын оятуға болады. Ойын жасағанда баланың жас ерекшелігі мен қызығушылықтарын ескерген дұрыс. 9-суретте көрсетілгендей қосымшадағы ойын өте қарапайым құрылған. Бала сандарды шатаспай табуы керек. Дұрыс тапқан сайын жаңа деңгейге көтеріледі. Бұл баланың жаңа ақпарат қабылдау арқылы миын жаттықтыруға ықпал етеді. Сөйлеу қабілеті нашар балалар көп жағдайда тыңдағанды ұнатады. Тыңдау арқылы бірінші миы жаттығып, ол тілінің дамуына әсер береді. Сол себепті, пайдаланушыға көңілді болуы үшін қазақша фондық әуен қойылды.

Firestore деректер қорында «Файлдарға» өтіп, әндер жүктеледі. Әнді қосышамен байланыстыру үшін URL мекен-жайы қажет. Ән файлының оң жақ шетін шертіп, URL мекен-жайын көшіру керек. Сілтемені басқаннан кейін ол автоматты түрде көшіріледі. Ең алдымен, файлға өтіп, оған сілтемені қойылады. `Oncreate()` әдісіндегі айнымалылар анықталады. Осыдан кейін `onCreate()` әдісінде Media player данасын орнату керек.



Сурет 9. Қосымшадағы ойын бөлімі және ойындарды таңдау беті

Android ОЖ (операциялық жүйе) үшін көптеген қосылатын кітапханалар болады, оларды мақсатына қарай келесі топтарға жіктеледі:

1. Үйлесімділік кітапханалары. Олар Android ОЖ-нің кейбір нұсқаларында, платформаның бұрынғы нұсқаларында пайда болған мүмкіндіктерді пайдалануға мүмкіндік береді. Себебі, API-нің жаңа нұсқалары осы нұсқаны қолдайтын құрылғылар кеңінен қолданылғаннан гөрі тезірек шығады.

2. Арнайы мақсаттағы кітапханалар. Ойындарды дамыту, әлеуметтік желілермен жұмыс істеу, статистика жинау және басқа жағдайларда қолданылады.

3. Қосымша мүмкіндіктер беретін кітапханалар. Бұл санатқа көптеген кітапханаларды жатқызуға болады. Бұған графикалық сурет кітапханалары, суреттермен жұмыс, өзгертілген басқару элементтері.

Кітапханалар жиналған және пайдалануға дайын `jar` файлдары түрінде жеткізілуі мүмкін. Кітапхананы қосу (файл `*.jar`) өте жеңіл орындалады. Жобада `libs` бумасын құру жеткілікті (`src` және `res` бумаларымен бірдей деңгейде) және кітапхана файлы сол жерге көшіру жеткілікті. Содан кейін оны жобаға `Project -> Properties` мәзірі арқылы қосу керек. Егер кітапхана бастапқы код түрінде ұсынылса, оны алдын-ала жинау керек. Жобаның түпкі бумасын оң жақ батырмамен басу керек `-> экспорттау: -> Java -> іске қосылатын файл -> іске қосу үшін`

Қойылған мақсатқа сай «Ainalaın» мобильді қосымшасы құрылды. Бұл зерттеу жұмыста Word2Vec әдісі негізінде сөйлеу қабілеті нашар балаларға арналған қазақ тілінің синонимдерінің сөздігі жасалды. Құрастырылған сөздікте 3000 сөз арнайы сөздер топтар бойынша топтастырылған. Бұл баланың тақырып бойынша сөздерді есте сақтауына жағдай жасайды. Әртүрлі дереккөздерден қазақша мәтіндер корпусын жинақталды, мәтінді алдын ала өңделді және сөздерді құру үшін Word2Vec әдісін оқытады. Модельдің өнімділігін әртүрлі сөз жұптары арасындағы косинус ұқсастығын өлшеу және модель қайтарған ең жоғары ұқсас сөздерді қолмен таңдалған синонимдер тізімімен салыстыру арқылы бағаланды.

АЛҒЫС

Бұл зерттеу Қазақстан Республикасы Ғылым және жоғары білім министрлігінің IRN AP 19577833 жобасының грантымен орындалды және қаржыландырылды.

Пайдаланылған дереккөздер тізімі

[1] Баранов А.А., Маслова О.И., Намазова-Баранова Л.С. // Онтогенез нейрокогнитивного развития детей и подростков // Вестник РАН. 2012, №(8), с.26–33.

[2] В Казахстане детей с речевыми проблемами становится все больше [Электронды ресурс]. – Қол жеткізу режимі: <https://astanatv.kz/ru/news/51629/>

[3] Дербисалова Г. С. Мультимедийные презентации как средство активизации познавательной деятельности младших школьников с нарушением интеллекта //Наука и перспективы. – 2019. – №. 3. – С. 52-58.

[4] Васильева В. С. Основные положения теории релевиной в современных научных исследованиях // Студенческий электронный журнал СтПИЖ. – 2021. – №. 2-2. – С. 88-90.

[5] Kumar R., Malik S., Gupta, S. (2019). Development of Hindi thesaurus using Word2Vec. Proceedings of the 3rd International Conference on Information Management and Machine Intelligence, 308-317. https://doi.org/10.1007/978-981-13-1801-4_28

[6] Mavridis T., Giannakidou D., Koutsombogera M. (2020). A Greek thesaurus using Word2Vec: A tool for language therapy. International Journal of Computational Linguistics and Applications, 11(2), 21-31. <https://doi.org/10.1515/ijcla-2020-0002>

[7] Serikbolova A., Shukeyeva M. (2019). Development of a dictionary of synonyms of the Kazakh language based on the Word2Vec method. Proceedings of the 2019 IEEE 9th International Conference on Consumer Electronics - Berlin (ICCE-Berlin), 1-5. <https://doi.org/10.1109/ICCE-Berlin.2019.8868532>

[8] Serikbolova A., Shukeyeva M. (2020). Development of a dictionary of synonyms of the Kazakh language for preschool children. Journal of Physics: Conference Series, 1605, 022031. <https://doi.org/10.1088/1742-6596/1605/2/022031>

[9] Serikbolova A., Shukeyeva M. (2021). Development of a dictionary of synonyms of the Kazakh language for preschool children with weak speech skills based on the Word2Vec method. Eurasian Journal of Educational Research, 21(91), 59-76. <https://doi.org/10.14689/ejer.2021.91.4>

[10] https://docs.google.com/spreadsheets/d/1PRw3i84thg8nDA9-NJ7u7bmSIUCkNLct/edit?usp=share_link&oid=114468051330637207467&rtpof=true&sd=true

[11] Что такое Android? Вот все, что вам нужно знать [Электронды ресурс]. – Қол жеткізу режимі: <https://www.androidauthority.com/what-is-android-328076/>

[12] Владимир Р. Облачная СУБД Firebase. – Кишинев: Tehnica UTM. – 2019. – С. 299.

[13] Zhang Q., Liu K. (2019). Research on English Synonym Dictionary Based on Word2Vec. Advances in Social Science, Education and Humanities Research, 326, 305-308. doi: 10.2991/ichss-19.2019.68

[14] Wu Y., Chen Q., Li W. (2018). Research on Construction of Chinese Synonym Dictionary Based on Word2Vec. Journal of Physics: Conference Series, 1057, 042009. doi: 10.1088/1742-6596/1057/4/042009

[15] Zhang Y., Cui Y., Liu X., Zhang J., Sun X. (2021). Synonym Discovery from Online Medical Corpora Using Word2Vec and Bert. Applied Sciences, 11(6), 2816. doi: 10.3390/app11062816

[16] Dehkharghani R. T., Vahdatnia M., Heydari P. (2020). Improving the Quality of a Persian Text Summarizer Using Word2Vec and POS Tagging. Journal of King Saud University-Computer and Information Sciences. doi:10.1016/j.jksuci.2020.05.003

[17] Tomas Mikolov, Kai Chen, Greg Corrado, Jeffrey Dean *Efficient Estimation of Word Representations in Vector Space*. <https://arxiv.org/pdf/1301.3781.pdf>

[18] Ganesa K. (2020, July 30). *Word2Vec: A comparison between CBOW, skipgram, skipgrams*. <https://kavita-ganesan.com/comparison-between-cbow-skipgram-subword/#.ZBHnBOxBwl8>

[19] *Архитектура Android-приложений. Часть III – основные части приложения [Электронды ресурс]*. <https://habr.com/ru/post/141201/>

[20] *Принципы переноса ранее разработанных приложений на смартфон [Электронды ресурс]*. Қол жеткізу режимі: https://intuit.ru/studies-/professional_skill_improvements /12785/ courses/1219/lecture/22490

References

[1] Baranov A.A., Maslova O.I., Namazova-Baranova L.S. (2012) *Ontogenez neyrokognitivnogo razvitiya detey i podrostkov [Ontogenesis of neurocognitive development of children and adolescents]*. Vestnik Rossijskoj akademii medicinskih nauk. №8, 26-33. (In Russian)

[2] *V Kazakhstane detey s rechevymi problemami stanovitsya vse bol'she [In Kazakhstan, there are more and more children with speech problems] [Elektrondy resource]*. About the emergency mode: <https://astanatv.kz/ru/news/51629/>. (In Russian)

[3] Derbisalova G. S. (2019) *Mul'timediynye prezentatsii kak sredstvo aktivizatsii poznavatel'noy deyatel'nosti mladshikh shkol'nikov s narusheniyem intellekta [Multimedia presentations as a means of enhancing the cognitive activity of primary schoolchildren with intellectual disabilities]*. Nauka i perspektivy. №3.52-58. (In Russian)

[4] Vasilyeva V. S. (2021) *Osnovnyye polozheniya teorii relevinoy tekhnologii v sovremennykh nauchnykh issledovaniyakh [Basic provisions of the Relevina theory in modern scientific research]*. Studencheskiy elektronnyy zhurnal StrIZH. № 2-2.88-90. (In Russian)

[5] Kumar R., Malik S., Gupta S. (2019). *Development of Hindi thesaurus using Word2Vec*. Proceedings of the 3rd International Conference on Information Management and Machine Intelligence, 308-317. doi: 10.1007/978-981-13-1801-4_28

[6] Mavridis, T., Giannakidou, D., & Koutsombogera, M. (2020). *A Greek thesaurus using Word2Vec: A tool for language therapy*. International Journal of Computational Linguistics and Applications, 11(2), 21-31. doi: 10.1515/ijcla-2020-0002

[7] Serikbolova, A., & Shukeyeva, M. (2019). *Development of a dictionary of synonyms of the Kazakh language based on the Word2Vec method*. Proceedings of the 2019 IEEE 9th International Conference on Consumer Electronics - Berlin (ICCE-Berlin), 1-5. doi: 10.1109/ICCE-Berlin.2019.8868532

[8] Serikbolova A., Shukeyeva M. (2020). *Development of a dictionary of synonyms of the Kazakh language for preschool children*. Journal of Physics: Conference Series, 1605, 022031. doi: 10.1088/1742-6596/1605/2/022031

[9] Serikbolova, A., Shukeyeva M. (2021). *Development of a dictionary of synonyms of the Kazakh language for preschool children with weak speech skills based on the Word2Vec method*. Eurasian Journal of Educational Research, 21(91), 59-76. doi: 10.14689/ejer.2021.91.4

[10] https://docs.google.com/spreadsheets/d/1PRw3i84thg8nDA9NJ7u7bmSIUCkNLct/edit?usp=share_link&ouid=114468051330637207467&rtopf=true&sd=true

[11] *Что такое Android? Вот все, что вам нужно знать [What is Android? Here's everything you need to know] [Electronds resource]*. How to choose the mode: <https://www.androidauthority.com/what-is-android-328076/>. (In Russian)

[12] Vladimir R. (2019) *Oblachnaya SUBD [Cloud DBMS Firebase]*. Chisinau: Tehnica UTM. 299. (In Russian)

[13] Zhang Q., Liu K. (2019). *Research on English Synonym Dictionary Based on Word2Vec*. Advances in Social Science, Education and Humanities Research, 326, 305-308. doi: 10.2991/icaahss-19.2019.68

[14] Wu Y., Chen Q., Li W. (2018). *Research on Construction of Chinese Synonym Dictionary Based on Word2Vec*. Journal of Physics: Conference Series, 1057, 042009. doi: 10.1088/1742-6596/1057/4/042009

[15] Zhang Y., Cui Y., Liu X., Zhang J., Sun X. (2021). *Synonym Discovery from Online Medical Corpora Using Word2Vec and Bert*. Applied Sciences, 11(6), 2816. DOI: 10.3390/app11062816

[16] Dehkharghani R. T., Vahdatnia M., Heydari P. (2020). *Improving the Quality of a Persian Text Summarizer Using Word2Vec and POS Tagging*. *Journal of King Saud University-Computer and Information Sciences*. DOI: 10.1016/j.jksuci.2020.05.003

[17] Tomas Mikolov, Kai Chen, Greg Corrado, Jeffrey Dean *Efficient Estimation of Word Representations in Vector Space*. <https://arxiv.org/pdf/1301.3781.pdf>

[18] Ganesan, K. (2020, July 30). *Word2Vec: A comparison between CBOW, skipgram, skipgrams*. <https://kavita-ganesan.com/comparison-between-cbow-skipgram-subword/#.ZBHnBOxBwI8>

[19] *Arkhitektura Android-prilozheniy. Chast' III – osnovnaya chast' prilozheniya* [Architecture of Android applications. Part III – main parts of the application] [Electronds resource]. About the emergency mode: <https://habr.com/ru/post/141201/> .(In Russian)

[20] *Printsipy perenosa raneye razrabotannykh prilozheniy na smartfon*. [Principles of transferring previously developed applications to a smartphone] [Elektrondy resource]. About the detailed regime: https://intuit.ru/studies-/professional_skill_improvements /12785/courses/1219/lecture/22490 .(In Russian)